

CSC-466

Analyzing Political Committee Election Donations

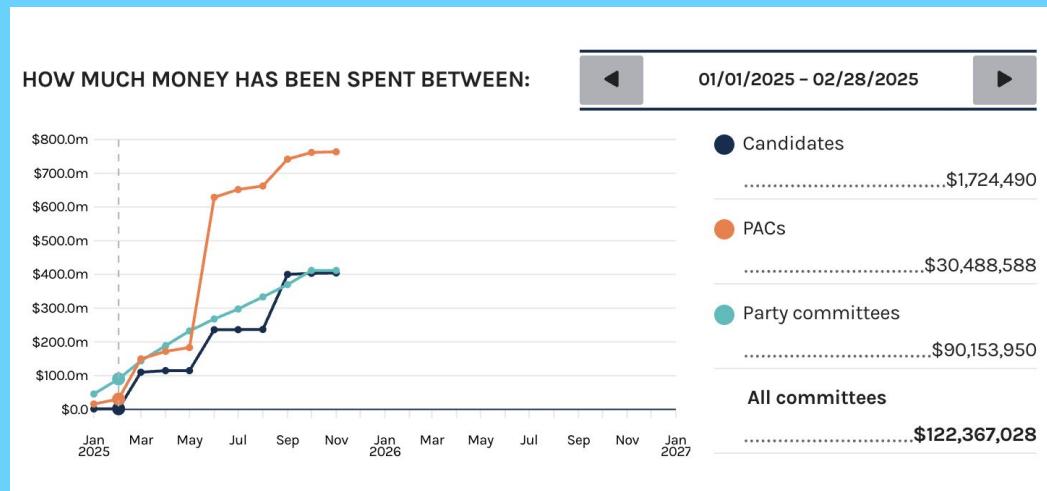
By Nipun Das

Why should you care?

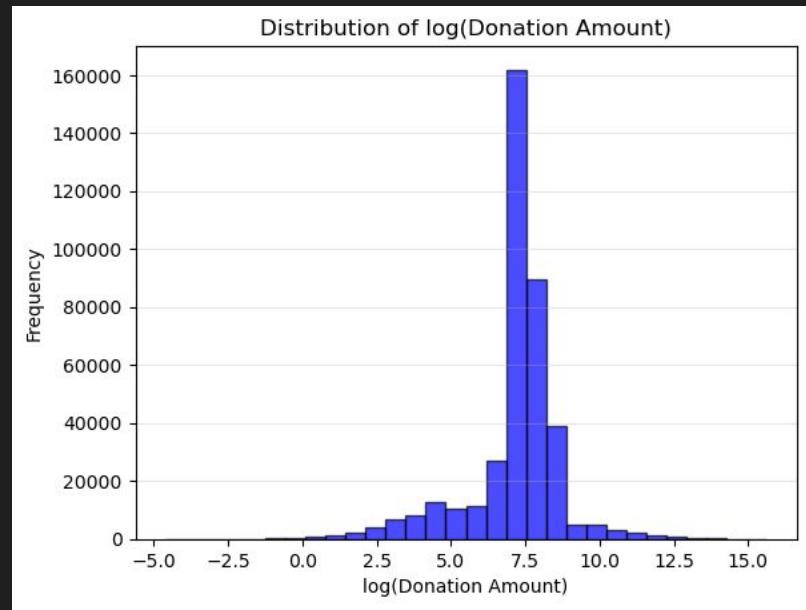
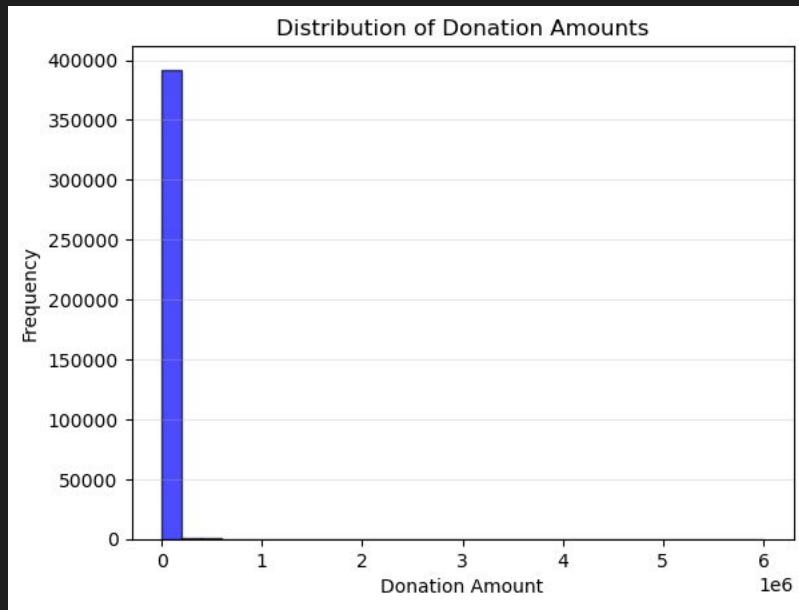
- We consistently see records broken with spending amounts in recent years, whether that's with each of the recent presidential cycles or with competitive congressional races (source: [OpenSecrets](#))
- News media coverage often focuses on press conferences, public statements, and political debates, but the increasing amount of money in politics may be more indicative of the policy decisions politicians make compared to generic politician rhetoric
- Analyzing political contributions data and understanding who funds your representatives provides us insights that aren't displayed publicly very often

Dataset Overview

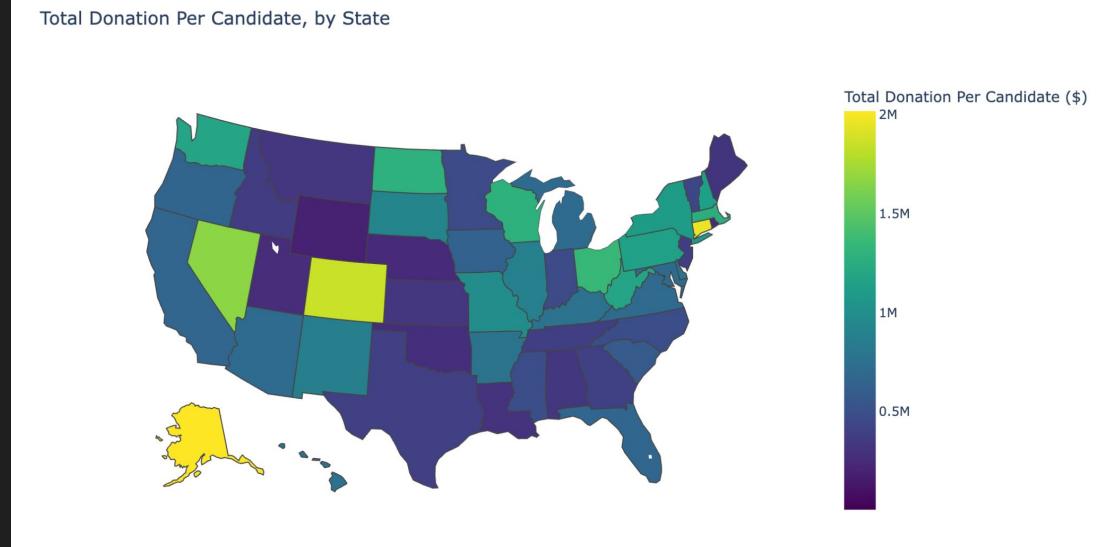
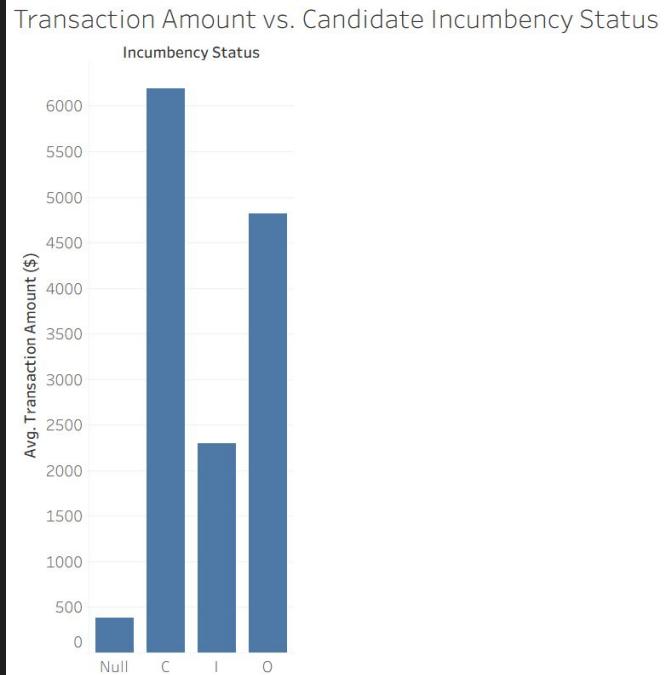
- Stanford DIME dataset (Database on Ideology, Money in Politics, and Elections)
- Includes contributions from individuals and various committees (PACs, Super PACs, etc.) to political candidates (state, federal, presidential elections)
- To narrow the scope (and use a manageable number of records), I chose to focus on donations from committees to federal congressional races in the 2010 election cycle
- Goal: predict which committees will donate to a candidate in the future, based on data midway through the election cycle
- Does not analyze spending, but notable that committees often spend large on behalf of candidates



Exploratory Data Analysis



Exploratory Data Analysis



Models

- Heuristic: for each committee, categorize them by party based on which party they donate the most to, then predict donors for each candidate as the committees that donate the most to
- Compared heuristic and random baselines to the following models:



Surprise Collaborative Filtering w/ SVD

Used min-max scaling on committee vectors to get “ratings” from donation amounts, then generate “recommendations” from interaction matrix with surprise (SVD).

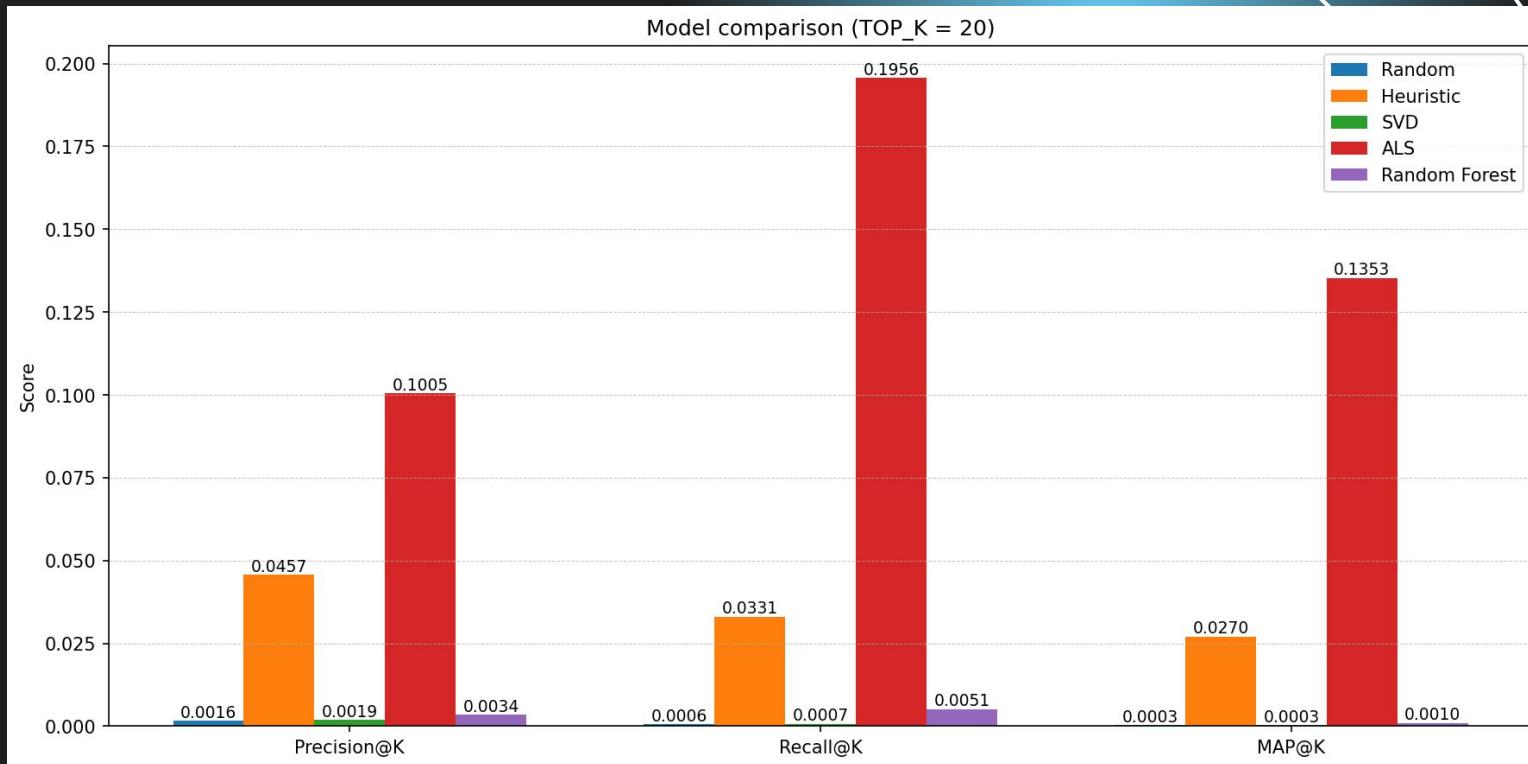
Collaborative Filtering w/ ALS

Alternating Least Squares (ALS) is a model-based collaborative filtering algorithm that extracts latent feature vectors and is often used with implicit ratings.

Random Forest

Using features about committees and candidates, predict a donation amount for each committee/candidate pair, and sort predicted donation amounts to get recommendations.

Results



Feature Engineering/Selection Challenges

- Temporal leakage with numerical features (ex. number of donations, total donations, latent features)
- Polling data could be useful but wasn't available
- Categorical features provide some signal but not sufficient
- Framing problem within one election cycle

Next Steps

(if I had more time)



Step 1

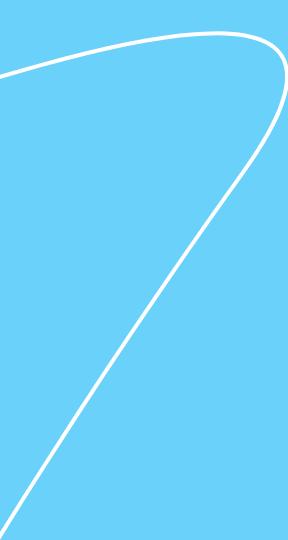
Explore seeing if more numerical features with useful signal can be obtained without temporal leakage, possibly through analyzing historical data or other datasets.

Step 2

Spend more time training a candidate/committee feature-based model, such as the existing random forest model or other models like XGBoost.

Step 3

Improve the ALS collaborative filtering model using a hybrid model with committee/candidate features.



Thank you!
